Journal of Biopharmaceutical Statistics, 18: 641–643, 2008

ISSN: 1054-3406 print/1520-5711 online
DOI: 10.1080/10543400802071337

Taylor & Francis
Taylor & Francis Group

COMMENTARY ON "ACCOUNTING FOR THE INTERIM SAFETY MONITORING OF AN ADVERSE EVENT UPON TERMINATION OF A CLINICAL TRIAL"

Jen-pei Liu
Department of Agronomy, Division of Biometry, National Taiwan University, Taipei, Taiwan and Division of Biostatistics and Bioinformatics, National Health Research Institutes, Zhunan, Taiwan

After more than 100 deaths caused by the Elixir Sulfanilamide disaster, in 1938, the U.S. Congress passed the Federal Food, Drug, and Cosmetic Act (FD&C Act) which for the first time in U.S. history required pharmaceutical companies to submit full reports of investigations regarding the safety of new drugs. Not until 1962, after the passage of the Leaver–Harris Amendment of the FD&C Act, was the U.S. Food and Drug Administration (FDA) authorized to require evidence of efficacy for approval of new drugs. Consequently, currently, for approval of a new drug the FDA requires that adequate and well-controlled clinical trials be conducted in humans to demonstrate substantial evidence of effectiveness and safety. In the evolution of regulations on approval of new drugs, therefore, safety of the drug was the primary issue and came before efficacy.

However, this may not be totally true for pharmaceutical companies when developing new drugs and for regulatory agencies when approving drugs, at least in design, conduct, and analysis of adequate and well-controlled clinical trials. The current paradigm for approval of a new drug is two-fold. First is to prove that the drug is efficacious. Second is to verify whether there is any excessive safety risk. Most adequate and well-controlled trials are sufficiently powered to evaluate efficacy either by superiority or by noninferiority. However, most so-called "adequate and well-controlled trials" are not powered to detect excessive risk by the following hypothesis:

$$H_0 : P_T \leq P_C \quad \text{vs.} \quad H_a : P_T > P_C, \tag{1}$$

where P_T and P_C are the proportions of patients with certain adverse events, for test and control groups, respectively.

In (1), we use $P_T \leq P_C$ simply to illustrate the concept; other parameters for evaluation of risk, such relative risk, odd ratio, or hazard ratios, can also be used.

Address correspondence to Jen-pei Liu, Department of Agronomy, Division of Biometry, National Taiwan University, Taipei, Taiwan and Division of Biostatistics and Bioinformatics, National Health Research Institutes, Zhunan, Taiwan; E-mail: JPliu@ntu.edu.tw

When rejecting the null hypothesis in (1), the evidence of excessive risk is substantial and if the adverse event is very serious, such as a cardiovascular event, then the new drug may not be approved. On the other hand, the new drug is approved if it can be shown to be efficacious and the null hypothesis in (1) is not rejected. However, failing to reject the null hypothesis in (1) does not imply that there is no evidence of excessive risk. In addition, as mentioned before, most adequate and well-controlled trials are not powered to detect excessive risk. Consequently, recently, some drugs, such as cerivastatin, Vioxx, and Celecoxib, just to name a few, were withdrawn from the market after approval.

Because the hypothesis in (1) cannot prove that the new drug is of no excessive risk, the noninferiority hypothesis that follows can be employed to prove that the risk of the new drug does not exceed a certain prespecified level, and hence to prove that the drug is of no excessive risk:

$$H_0 : P_T - P_C \geq \delta_0 \quad \text{vs.} \quad H_a : P_T - P_C < \delta_0, \tag{2}$$

where δ_0 is some prespecified allowable limit for the risk.

Based on the concept of proving no excessive risk, Dallas (2008) in this issue proposes a procedure for interim safety monitoring of a rare serious event upon termination of a clinical trial. Under the assumption of a Poisson distribution for rare events, the authors propose methods for the adjusted p-value, point estimate, and confidence limits upon termination of the trial. However, it should be noted that the approaches for testing the hypotheses in (1) and (2) are totally different. For the hypothesis in (1), one would reject the null hypothesis at the α significance level and conclude existence of excessive risk if the $(1 - \alpha)\%$ lower confidence limit for $P_T - P_C$ is greater than zero. On the other hand, the null hypothesis in (2) is rejected at the a significance level, and one can conclude no excessive risk if the $(1 - \alpha)\%$ upper confidence limit for $P_T - P_C$ is less than the prespecified allowable limit, δ_0. Therefore, the evaluation of probability coverage and the width of the confidence interval noted in the paper is for the hypothesis of equal risk. Hence it is not directly relevant either to the hypothesis in (1) for detection of an excessive risk or to the hypothesis in (2) for proof of no excessive risk. In addition, for proving that a new drug is safe, the simulation study should investigate the empirical size of the proposed procedure for evaluation of no excessive risk, and this is much more important than the probability coverage and width of the confidence intervals.

The other issue for the hypothesis in (2) is the definition of excessive risk and determination of the prespecified allowable upper limit for the risk under consideration. This problem needs to be resolved before the concept of proving no excessive risk, based on hypothesis in (2), can be implemented in clinical practice and can be accepted for regulatory agencies. However, similar to the equivalence limits or noninferiority margins used for evaluation of efficacy, the prespecified allowable upper limit can be determined by the size of the target patient population, type and seriousness of the disease, and prevalence and severity of the adverse event. However, an upper allowable limit of 8 for excessive risk, suggested in the example of the paper, is exceedingly too liberal. A reasonable range for the upper allowable limit for excessive risk may be from 1.5 to 3.0.

In summary, the current paradigm of evaluation of safety by detection of excessive risk is not adequate to prove that a new drug is safe. Under the concept

of risk management, to prove that a drug is safe is to prove that the drug is of no excessive risk. It turns out, this means proving that the risk of the drug does not exceed some prespecified allowable upper limit. This is the new concept brought by the paper. However, the statistical methodology for this new concept is already available, for example in Liu et al. (2005). Neverthless, the philosophy for evaluation of safety behind the two approaches will continue to debate for years to come.

REFERENCES

Dallas, M. J. (2008). Accounting for the interim safety monitoring of an adverse event upon termination of a clinical trial. *Journal of Biopharmaceutical Statistics* 18(4):631–638.

Liu, J. P., Fan, H. S., Ma, M. C. (2005). Tests for equivalence based on odds ratio for matched-pair design. *Journal of Biopharmaceutical Statistics* 15:889–901.